DETECTING CHANGES IN MEAN LEVELS WITH ATHEORETICAL REGRESSION TREES

Carmela Cappelli[†] **and Marco Reale** *

* *Department of Mathematics and Statistics*
University of Canterbury
Private Bag 4800
Christchurch, New Zealand

Report Number: UCDMS2005/2 MARCH 2005

Keywords: Multiple structural breaks at unknown times, time series, Fisher algorithm, Bai and Perron procedure, least square regression trees, non-parametric testing, hierarchical models.

[†] *Department of Statistical Sciences, University of Naples Federico II, Naples, Italy*

Detecting Changes in Mean Levels with Atheoretical Regression Trees

Carmela Cappelli and Marco Reale

March 16, 2005

Abstract

In this paper the study of long term changes in lake mean water levels is faced as a problem of detecting multiple structural breaks in the mean occurring at unknown dates. We propose a non parametric approach that exploits, in the framework of least squares regression trees, the contiguity property of the Fisher method proposed for grouping a single real variable. The proposed approach is applied to study the changes in mean water levels of Lake Michigan-Huron.

1 Introduction

The study of lake water levels is a relevant issue in hydrology and in general in the geosciences because their fluctuations have climatological, ecological and economic consequences. Indeed, water level fluctuations can affect shoreline erosion, aquatic ecosystem, riparian interests, commercial navigation and several other water dependent facilities such as drinking water intakes and hydroelectric power generation. Thus, understanding the behavior of lake levels is a crucial component of any sensible strategy for coping with levels fluctuations.

In this paper we focus on long term variations in the mean water levels studying the case of Lake Michigan-Huron. The objective is to verify whether the lake mean water levels show a long-term lowering that is a main concern.

Long term changes in mean water levels can be meant as structural breaks whose detection is a challenging task and it has attracted lot of attention both in the statistic and econometric literature (for a review see Hansen, 2001).

Recently, Bai and Perron (1998, 2003) have proposed a non-hierarchical procedure to detect multiple structural breaks occurring at unknown date that makes use of a dynamic programming approach that can be traced back to the Fisher's method of exact optimization (Fisher, 1958) proposed for grouping n elements into G mutually exclusive and exhaustive subsets having maximum homogeneity, i.e. minimizing the within-groups sum of squares.

Fisher considers *contiguous partitions*, i.e. his method is designed for situations in which the data points are ordered and groups consist of intervals of data. He deals with two subclass of problems: the *unrestricted* case when the observations are ordered according to their numerical values, and the *restricted* one when an *a priori* ordering is given.
Lake water levels data belong to the second case as the ordering is provided by the time and observations are not exchangeable. In this case seeking the minimum within sum of square partition corresponds to segment the series of lake water levels into subperiods that contrast with each other being stationary in mean.
The idea of contiguous partitions can be exploited within the framework of regression trees to achieve the same goal, i.e., to locate the break dates.
Given a continuous response variable Y and a set of p predictors $X_1, X_2, \ldots, X_p$, regression trees model the relationship between the response and the covariates employing a recursive partitioning approach that results into a partition of Y based upon the values of the predictor variables. Our procedure makes use of an artificial covariate (so that $p = 1$) that is an arbitrary strictly ascending (or descending) sequence of numbers. Thus we call the procedure Atheorethical Regression Trees (so forth denoted by ART) because it is theory-free being the covariate not a predictor variable but rather a counter.
In what follows we will show that the use of such covariate in least square regression trees (Morgan and Sonquist, 1963; Breiman *et al.*, 1984) resorts to a recursive application of the Fisher's method to a problem of $G = 2$ subperiods producing an open nested partition (Boros, 1996) i.e., a hierarchical structure in the form of a binary tree whose split points correspond to candidates break dates. The final set of break dates and corresponding homogeneous subperiods can result either from automatic procedure such as pruning or from the subjective choice of the applied scientist based on *a priori* knowledge. Thus the procedure is data driven since the number of breaks and the times of occurrence are unknown and they are uncovered by the procedure itself.
The remainder of paper is organized as follows. Section 2 casts the problem of identifying long term changes in lake mean water levels in terms of detection of structural breaks in the mean. In section 3 we introduce regression trees and the Fisher's algorithm of exact optimization in order to show the connections existing between them that result in the ART method. In section 4 various criteria to select the final set of break dates and the corresponding homogeneous subperiods are discussed. In section 5 the proposed approach is applied to the study of changes in the mean water levels of lake Michigan-Huron. A comparison of our procedure with Bai and Perron's is presented in section 6. Final remarks follow in section 7.

2 Problem restatement

Lake water levels can undergo *short term changes* due for example to storms or ice jams, *seasonal changes* in response to the interannual variability of precipitations and evaporations and *long term changes* due to prolonged climatic variations as well as to human intervention.

Although short term variations can cause severe damages, the long term ones have the highest impact on the environment, the ecosystem and the human settlements.
With the aim of studying the year to year changes, we consider yearly data and we assume that there are intervals of years when the lake mean water levels remained fairly constant, shifting from one interval to the other. Thus, the following mean-shift model can be considered:

$$y_t = \mu_g + \epsilon_t, \quad g = 1, \dots, G-1, \quad t = T_{g-1}+1, \dots, T_g, \tag{1}$$

where $G-1$ is number of break dates (mean water level changes), y_t the lake water level (yearly mean level) and ϵ_t is the error term at time t (we adopt the common convention that $T_0 = 0$ and $T_G = T$ where T is the length of the series). This model has been employed by Bai and Perron (2003) to study multiple structural breaks in the mean occurring at unknown dates. The problem resorts to estimate the set of break dates $\{1, \dots, g, \dots, G-1\}$ that define a partition of the series $P(n, G) = \{I_1, \dots, I_g, \dots, I_G\}$, into homogeneous segments such that $\mu_g \neq \mu_{g+1}$.
Note that the segmentation process, along with the detection of structural breaks, is useful for two more issues in time series analysis. First, in the context of forecasting a time series it is sensible to base the forecasts on a model estimated on a recent segment of the series instead of using the entire series and this is especially true for time series covering extended period for which the specification of a single model can be inadequate. Second, the segmentation process might isolate short intervals between longer ones revealing the presence of outliers and thus suggesting the need for adjusting the data. Third, the presence of structural breaks reveals a behavior of the time series that could otherwise be misunderstood and modeled inadequately. In particular such a presence may lead to an erroneous identification of an integrated or fractionally integrated process (Perron, 1989; Hidalgo and Robinson, 1996; Granger and Hyung, 2004).

3 Atheoretical regression trees

Let (Y, X) be a random vector, with $Y \in R$ and $X \in R^p$, regression trees seek a function $f(X)$, for predicting the response variable Y given values of the predictor variables X.
As error function of the predictor $f(X)$, the mean squared error $E(Y - f(X))^2$ is commonly employed. Use of this measure leads to *least squares regression trees* (LSRT) in which $f(X)$ is the conditional expectation $E(Y|X = x)$. Thus, LSRT fit to each tree node the group mean, i.e., the mean of the Y's values falling into the node, because this represents the optimal (or Bayes) prediction minimizing the mean squared error (for complete discussion on this issue see Breiman *et* al., 1984 ch.9).
Based on a training set $(y_i, x_{i1}, \dots, x_{ip})_{i=1}^n$, the algorithm proceeds by recursively splitting the data into two subsets. Any split is a binary question of the form: "Is $x_j \in A$", so that, in case of a numeric predictor variable, the set of possible splits includes all questions: Is $x_j \leq c$?, for c ranging over the domain of x_j. The split induces a partition of the observations y_i: the left descendant nodes h_l satisfying $\{x_{ij} \leq c\}$ and the right descendant

node h_r satisfying $\{x_{ij} > c\}$.
Thus, at any node h the algorithm selects the split s which maximally distinguishes the response variable in the left and the right descendant nodes providing the highest reduction in deviance

$$SS(h) - [SS(h_l) + SS(h_r)] \tag{2}$$

where $SS(h) = \sum_{y_i \in h}(y_i - \bar{y}(h))^2$, $(i = 1, \ldots, n)$, is the sum of squares for node h, and $SS(h_l)$ and $SS(h_r)$ are the sums of squares for the left and right descendants, respectively. As h_l and h_r are an exhaustive partition of h, $SS(h)$ represents the total sum of squares $TSS_y(h)$ and $SS(h_l) + SS(h_r)$ the within-group sum of squares $WSS_{y|s}(h)$. Therefore the splitting criterion stated in (2) is equivalent to maximize the between-groups sum of squares $BSS_{y|s}(h)$ given by

$$BSS_{y|s}(h) = \sum_{q \in \{l,r\}} n(h_q)(\bar{y}(h_q) - \bar{y}(h))^2 \tag{3}$$

where $n(h)$ denotes the number of y values in node h and $n(h_l)$ and $n(h_r)$ the part that go to left and right, respectively.
Note that, since $\bar{y}(h) = \frac{n(h_l)\bar{y}(h_l) + n(h_r)\bar{y}(h_r)}{n(h)}$, (3) can be rewritten as

$$BSS_{y|s}(h) = \frac{n(h_l)n(h_r)}{n(h)}(\bar{y}(h_l) - \bar{y}(h_r))^2,$$

this formulation shows that in LSRT the splitting criterion separates subgroups of y's values whose means are as far as possible.
Once the binary partition of a node is found, the splitting process is applied separately to each subgroup, and so on recursively until the subgroups either reach a minimum size or no improvement of the criterion can be achieved. The resultant tree usually is overly large so that a pruning method is applied to trim it back. Alternatively a stopping rule can be employed, i.e., growing stops unless a condition, usually set in terms of node size or node accuracy, is met.
Minimizing the within-group sum of squares or, equivalently, maximizing the between group sum of squares is a natural clustering criterion for partitioning a single real variable (Everitt *et al.*, 2001).
This is the case of the Fisher's algorithm of exact optimization where the concept of *contiguous partitions* is introduced.
Let i, i' and i'' be three data points such that $i < i' < i''$; according to Fisher a partition is said to be contiguous if it consists of groups that satisfy the following condition: if i and i'' are assigned to the same class then i' must be also assigned to that class.
For ordered data only contiguous partitions require to be considered to detect the optimal one minimizing the within-group sum of squares. In the case of restricted problems contiguity is defined with respect to the *a priori* ordering. Thus, in the case of lakes water levels data the contiguity applies to time, i.e. only consecutive intervals in terms of the ordering specified by time are admissible.

Fisher demonstrates that least square partitions, i.e. partitions minimizing the within group sum of squares, are contiguous and his algorithm applied to ordered data points finds the exactly optimal partition into G groups.
The number of possible contiguous partitions of n (whenever) ordered objects into g groups makes a global search unfeasible but Fisher shows that the number of computations can be substantially reduced by exploiting the additivity property of the sum of squares criterion by means of dynamic programming algorithm (Bellman and Dreyfus, 1962) that allows to deal with the problem of finding the optimal partition into G groups making use of the results obtained while dealing with the problem of $G-1$ groups. This efficient algorithm is employed by Bai and Perron (2003) to find the global minimizer of the sum of squares, but, despite the computational saving it cannot deal with any value of n and G.
We have found that the concept of contiguity introduced by Fisher can be naturally exploited in the framework of least square regression trees. To this aim let k be an arbitrary ascending (or descending) sequence of completely ordered numbers, for sake of simplicity take $k = 1, 2, \dots, i, \dots, n$.
The use of such sequence as covariate into least square regression trees resorts to create and check at any node h all the $n(h)-1$ possible binary contiguous partitions of the $y_i \in h$. Thus, the covariate is not to be considered as a predictor variable but rather as an auxiliary variable, a counter, that allows the tree algorithm to try all admissible splits. In this respect, the procedure is theory-free and for this reason we call it Atheorethical Regression Trees (ART).
The contiguity property ensures that for any node h the best split lays in k (or in its subintervals after the split of the root has taken place) and the tree algorithm, based on splitting criterion (3), is forced to identify it.
Note that in the original Fisher's method as well as in the Bai and Perron procedure, optimal partitions for different values of G need not to be hierarchically nested. ART method is based on a binary search algorithm and as splitting goes on, the previous partitions are fixed. Thus, after several splits there's no guarantee that the global optimum is reached i.e., that the absolute minimum within groups sum of square partition is generated. It is so only after a single split but, as noticed in Gordon (1973), for many sets of data binary divisions represents a reasonable approximation providing good partitions (see also Edwards and Cavalli-Sforza, 1965).
In the case of time series data Hartigan (1975) provides an excellent justification in favor of the (faster) binary division algorithm: suppose that the observed time series consists of G segments within each of which the values are constant, i.e. model (1) becomes a piecewise constant model with $\epsilon_t = 0$. Then, there is a partition into G segments for which the within sum of squares is zero and it will be identified by a sequential splitting algorithm as the one in ART.
In other words, if the data have a hierarchical structure then ART will find the overall optimum, otherwise it provides a suboptimal solution for which, being the partitions contiguous, misplacements can occur only on the boundaries. As discussed in Hansen (2001), although structural breaks are treated as immediate, it is more reasonable to think that they take a period of time to become effective, thus misplacements on the boundaries are

not a concern.
Given that the global search algorithm requires $O(Gn^2)$ steps, whereas ART, at any tree node requires $O(n(h))$ steps to identify the best split, suboptimality does not appear a high price to pay to obtain full feasibility and indeed, in the application we will show that the partitions provided by ART are comparable to those obtained by the global search procedure.
Note that ART is fully nonparametric, i.e., the tree growing process does not require any distributional assumption. Nevertheless, if estimation is not the sole concern and one want to test for structural breaks or model the observations in the segments, it can be appropriate to consider restrictions on the possible values of the break points as suggested by Bai and Perron. Indeed, side conditions on the reduction in deviance and/or on the length of the subperiods are easily handled within the tree growing recursive partitioning approach of ART.

4 Getting the right number of subperiods

A drawback of partitioning methods such as Fisher's is that they produce a single partition for a prespecified value of G and, in general, it is advisable to produce and compare more partitions by varying G. In the case of ART this is not a concern because it produces a hierarchy, that is indeed an advantage because the inspection of the tree diagram allows an insight into the partitioning process also providing an ordering of the break dates based on their position into the tree and the reduction of the error function achieved.
On the other hand, as any recursive partitioning method, it is based on a *divide and conquer* algorithm that tends to grow a large initial tree T_{max}. Any subtree of T_{max} provides a possible set of break dates represented by the splitting points along the artificial covariate k which specify a partition of the series into segments. The identification of a final partition requires the space of all possible subtrees of T_{max} to be searched. This space is typically too vast; in the CART method a backward procedure is suggested to generate a collection of candidate partitions.
To this aim a cost function based on the concept of *cost-complexity measure* is introduced. Let A_h be the branch rooted at node h and denote by $\bar{A}_h$, the set of terminal nodes of the branch whose cardinality $|\bar{A}_h|$, is the cost-complexity measure for node h and its branches are defined as:

$$R_\alpha(h) = R(h) + \alpha,$$

$$R_\alpha(A_h) = R(A_h) + \alpha|\bar{A}_h|,$$

where $R(h) = SS(h)$ and $R(A_h) = \sum_{h \in \bar{A}_h} SS(h)$.
The two measures equals if:

$$\alpha_h = \frac{R(h) - R(A_h)}{|\bar{A}_h| - 1}, \tag{4}$$

From (4) we see that α_h gives the reduction in deviance for any additional split in the given branch. Therefore, when the two measures coincide, there is no point in retaining

the branch which increases the size without improving the accuracy.
Discarding at each step the subtree associated with the minimum value of α_h produces a sequence of candidate (finer and finer) partitions and break-date sets to which restrict the search of the final partition which corresponds, hopefully, to the actual number of distinct homogeneous subperiods present in the series. To this aim different possibilities are available, namely cross validation (CV), as proposed in the CART book, and model selection criteria.
Concerning the latter, Cappelli and Siciliano (1998) have employed the AIC (Akaike, 1973) in the context of classification trees to select the final tree among pruned trees produced by different pruning methods. Many other new criteria have appeared in the last decades and since none of them can be unconditionally recommended as a default procedure we consider here the most popular ones, namely, AIC, BIC (Schwarz, 1978), AICC (Hurvich and Tsai, 1989) and RIC (Shi and Tsai, 2002).
These criteria, under a Gaussian distribution assumption, revolve around a penalized sum of squares criterion where the penalty is a function of the model parameters that, in the case of tree structures, is given by the number of terminal nodes. Thus, these criteria are similar in spirit to the cost complexity measure. For explicit formulation and discussion on the use of model selection criteria within regression trees see Su *et al.* (2003), where they mainly considered the Gaussian context but also tested these selection criteria under different distributional behavior. As it is desirable to keep the ART methodology non parametric, we have tested extensively the robustness of the above mentioned information criteria for non Gaussian distributions with Monte Carlo simulations. All the criteria were found quite robust, especially the BIC and the RIC, confirming the findings of Su *et al.*.
It's worth noticing that trees allow a great deal of flexibility and that manual pruning based on subjective choices of the analyst can be preferred to automatic procedure as pointed out in Zhang and Singer (1998) where it is also discussed an alternative pruning method.

5 Application to Lake Michigan-Huron data

We apply the ART procedure to analyze the changes in water levels of the Lake Michigan-Huron. Whereas Lake Michigan is wholly within the borders of the US, Lake Huron is shared with Canada but the two lakes, although geographically distinct, are hydrologically inseparable and they represent one of largest body of water worldwide. Parts of the shorelines are urbanized and industrialized, others are intensively farmed and some others are wild and represent the ideal habitat of rare plants and animals. All these reasons explain the importance of Lake Michigan-Huron and the attention that it is payed by governmental agencies of US and Canada, by the scientific community and by the media.
The data-set reports the lake yearly mean water levels over a time span of 140 years running from 1860 (the first year of systematic water level recording) to 1999. The series has been partitioned by ART setting the minimum length of each segment equal to 5 years, a choice suggested by hydrological science, producing a maximal tree of 10 terminal nodes that is depicted in figure 1. The break dates can be retrieved simply adding the split point

to the initial year of the series. The inspection of this tree might suggest to the scientist

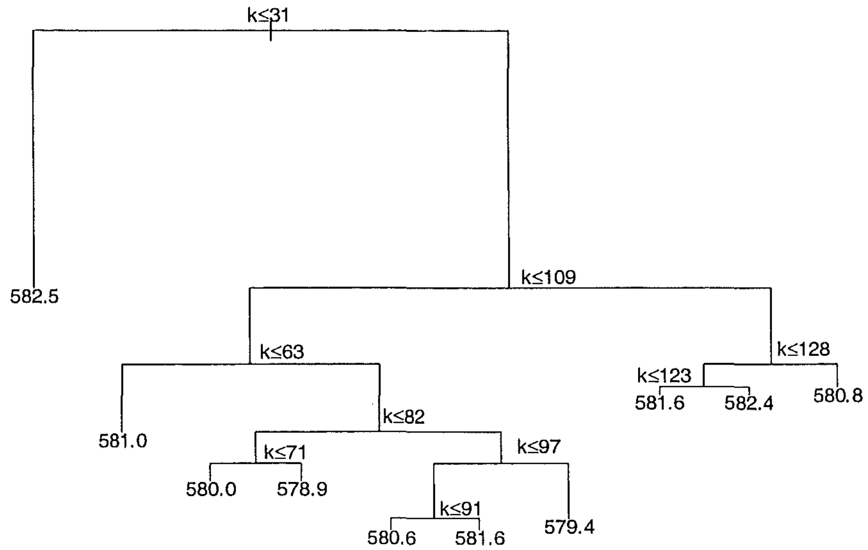

Figure 1: Large initial tree grown by ART. Values above the internal nodes report the split points, values beneath the terminal nodes indicate the mean water levels (feet) of the corresponding subperiods.

that one partition is preferable to another. Here, we consider the above described pruning method to create a sequence of subtrees and then, in order to select the final set of break dates identifying periods of fairly constant mean water level, we have applied CV, AIC, AICC, RIC and BIC.
Figure 2 reports the values of the cross validated error and of the various criteria against the number of terminal nodes of the trees in the sequence. Since the best partition is the one minimizing the values of the model selection criteria, from the graphs we see that CV would select the 8 segments partitions, AIC and AICC would the 10 segments partition (i.e., the partition corresponding to the maximal tree depicted in figure 1) and the BIC and RIC the one with 9 segments. Figure 3 depicts the corresponding partitions.
The graphs show that Lake Michigan-Huron has experienced a long period of stable high water levels till the last decade of the 19th century when they dropped to lower but still stable levels lasting for about more thirty years. Starting from the mid 1920's lake levels have began to alternate irregularly between shorter periods of lower and higher levels. Low unstable water levels have been recorded over the years 1923-1931 and the dry hot years of the 1930's; extreme low water levels occurred in the mid 1960's with the record low in 1964, likely in response to a navigation channel dredged in the St. Clair River that is the main outflow for Lake Michigan-Huron. A long period of sustained high levels has been recorded during the 1970's and the 1980's with the record high level of 1986.

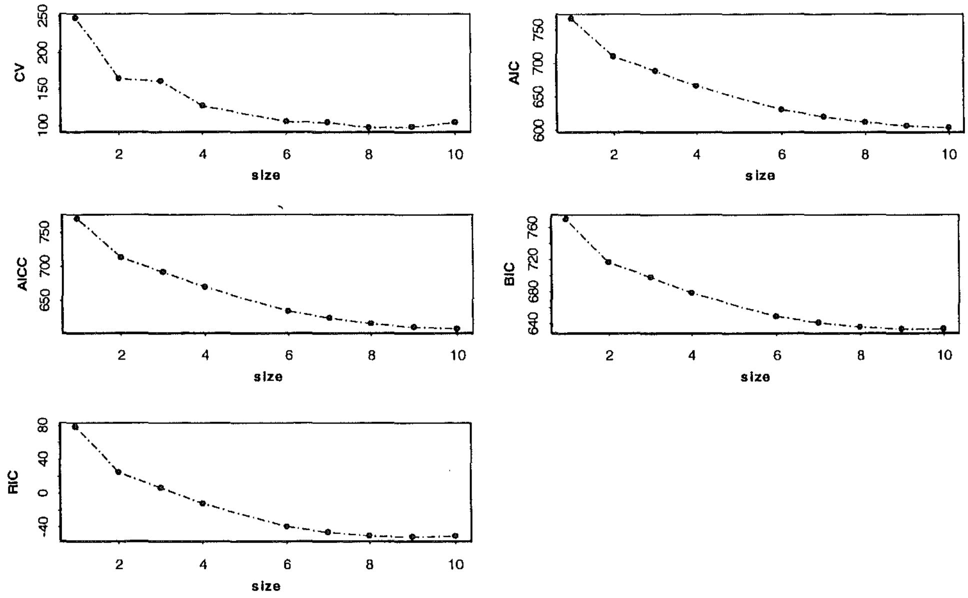

Figure 2: Values of the cross validated error and model selection criteria against the number of segments.

As a matter of fact, there are concerns that global warming and the so called greenhouse effect might cause an irreversible lowering of lake water levels. Our study allows an insights into the phenomenon showing that up to now no long term lowering can be identified and that lake mean water levels are characterized rather by a high instability over the years, alternating higher and lower phases.
Concerns for the future are certainly justified based on projected scenarios of climate shift (see for example Morsch and Quinn, 1996) towards warmer temperatures and decreased precipitations.

6 Comparison with the global search procedure

In order to compare our procedure with the global minimizer of the sum of squares, we have applied to the Lake Michigan-Huron data the procedure of Bai and Perron as implemented in the R package `strucchange` (Zeileis *et* al, 2002) being R (version 2.0.1, http://www.r-project.org) the environment in which we have implemented ART. Table 1 gives the partitions for G ranging from 2 to 10 with the indication of the corresponding within-sum of squares. The partitions are obtained setting the minimum length of any segment equals to 5 as in the application of ART. As expected these partitions are not completely nested, but, comparing the breakpoints identifying the partitions with the those in the tree grown

G	$SS(P_G)$	P_G								
2	155.76	31								
3	131.76	31	109							
4	110.14	30	63	109						
5	100.25	30	63	82	109					
6	82.53	30	63	83	97	108				
7	75.48	30	63	83	97	109	128			
8	70.31	30	63	71	82	97	109	128		
9	66.13	30	63	71	82	91	96	109	128	
10	63.69	30	62	67	72	82	91	96	109	128

Table 1: Partitions with global search procedure.

by ART (see figure 1) we see that the tree contains the same partitions with most of the break points being identical and single misplacements occurring on the boundaries. These misplacements cause the within sum of squares to be higher in the case of ART, indeed, for $G = 8, 9$ the within sum of squares of the ART partitions is 70.78 and 67.24 respectively, thus, the loss is not remarkable. The only partition that is not in the tree is for $G = 10$, in fact, the nested nature of ART partitions, together with the constraint on the segments length makes impossible to have a further break between observations 63 and 71 as in the global search procedure. For this reason the global search algorithm tends to create a much larger number of candidate partitions with respect to ART.
A second aspect to be compared concerns the computational costs associated with the two procedures. On an Intel Pentium 4 1.8GHz the CPU time of executing the global search for the Huron-Michigan analysis is 6.11s against 0.03s of ART, i.e. comparable results can be obtained in a much shorter time. Indeed, when it comes to long series, the global search procedure is unfeasible unless the minimum length of the segments is set to a value that strongly reduces the number of admissible partitions to be enumerated and evaluated.

7 Concluding remarks

We have proposed a data driven non parametric procedure for detecting multiple structural breaks in the mean occurring at unknown dates. The method, called ART, exploits the concept of contiguous partitions introduced by Fisher (1958) within the framework of least squares regression trees resorting to a sequential use of his algorithm of exact optimization. Although our procedure does not find the global minimum, its results are comparable to those obtainable applying Bai and Perron's procedure, coinciding for most of the break dates. The main advantages of the proposed approach are:

1. *simplicity* - it can be easily implemented or run with packages implementing routines to grow and prune least square regression trees;

2. *feasibility* - it can be used to find the least squares partition of an ordered sequence with no limitations either the number of observations or the number of segments;

3. *visualization*- it results in a tree diagram that allows to visualize at once the partitioning process, allowing for imputation of *a priori* knowledge in a nested hierarchy.

In this paper the method has been applied to explore changes in lake mean water levels that are known to cause social, economic and ecosystem disruption. In particular, declining lake levels are of concern to the Great lakes community but the application to Lake Michigan Huron data has shown that, despite the widespread concern no sign of a lowering trend are present up to now.

References

Akaike, H. (1970), Information theory and an extension of the maximum likelihood principle. In: *2nd International Symposium on Information Theory*, Petrov B.N. and Csaki, F. eds, 267-281, Akademia Kadio, Budapest.

Bai, J. and Perron, P. (1998), Estimating and testing linear models with multiple structural changes. *Econometrica*, 66, 47-78.

Bai, J. and Perron, P. (2003), Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, 18, 1-22.

Bellmann, R.E. and Dreyfus, S.E. (1962). *Applied Dynamic programming*, Princeton University Press.

Boros, E. and Hwang F. K.(1996), On the optimality of nested partitions, *SIAM Journal on Optimization*, 6, 1153-1162.

Breiman, L., Friedman J.H., Olshen R.A., and Stone, C.J. (1984). *Classification and Regression Trees*, Wadsworth & Brooks, Monterey (CA).

Cappelli, C. and Siciliano, R., (1998), Strategies for choosing the best decision tree. In: Ferndez Aguirre K. e Morineau A. eds, *Analyse Multidimensionnelle des Donnes*, 101-110, CERESTA-CISIA, Saint Mand (Fr).

Edwards, A.W.F. and Cavalli-Sforza, L.L. (1965), A method for Cluster analysis. *Biometrics*, 21, 362-375.

Everitt, B., Landau, S. and Leese M. (2001). *Cluster Analysis*, Arnold, London.

Fisher W.D. (1958), On grouping for maximum homogeneity. *Journal of the American Statistical Association*, 53, 789-798.

Gordon, A. D.(1973), Classification in presence of constrains, *Biometrics*, 29, 821-827.

Granger, C.W.J. and Hyung, N. (2004), Occasional structural breaks and long memory with an application to the *S&P 500* absolute stock returns. *Journal of Empirical Finance*, 11, 399-421.

Hansen, B.E. (2001), The new econometrics of structural change: dating breaks in the U.S. labor productivity. *Journal of Economic Perspectives*, 15, 117-128.

Hartigan, J.A. (1975). *Clustering Algorithms*, John Wiley and Sons, New York.

Hidalgo, J. and Robinson, P.M. (1996), Testing for structural change in a long-memory environment. *Journal of Econometrics*, 70, 159-174.

Hurvich, C. and Tsai, C.L. (1989), Regression and time series model selection in small samples. *Biometrics*, 76, 297-307.

Morgan, J.N. and Sonquist, J.A. (1963), Problems in the analysis of survey data and a proposal. *Journal of American Statistical Association*, 58, 415-434.

Mortsch, L. D. and Quinn, F. H. (1996), Climate change scenarios for Great Lakes Basin ecosystem studies. *Limnology and Oceanography* 41, 903-911.

Perron P. (1989), The great crash, the oil price shock and the unit root hypothesis. *Econometrica*, 57, 1361-1401.

Schwarz, G. (1978), Estimating the dimension of a model. *Annals of Statistics*, 6, 461-464.

Shi, P. and Tsai, C.L. (2002), Regression model selection- a residual likelihood approach. *Journal of the Royal Statistics Society*, Series B, 64, 237.252.

Su, X.G., Wang, M., and Fan, J.J (2004), Maximum Likelihood Regression Trees. *Journal of Computational and Graphical Statistics*, 13, 586-598.

Zeileis, A., Leisch, K. Hornik, K. and Kleiber, C. (2002), `strucchange`: an R package for testing for structural breaks in liner regression models. *Journal of Statistical Software*, 7, 1-38. URL `http://www.jstatsoft.org/v07/i02/`.

Zhang, H. and Singer, B.(1998). *Recursive partitioning in the health science.* Springer-Verlag, New York.

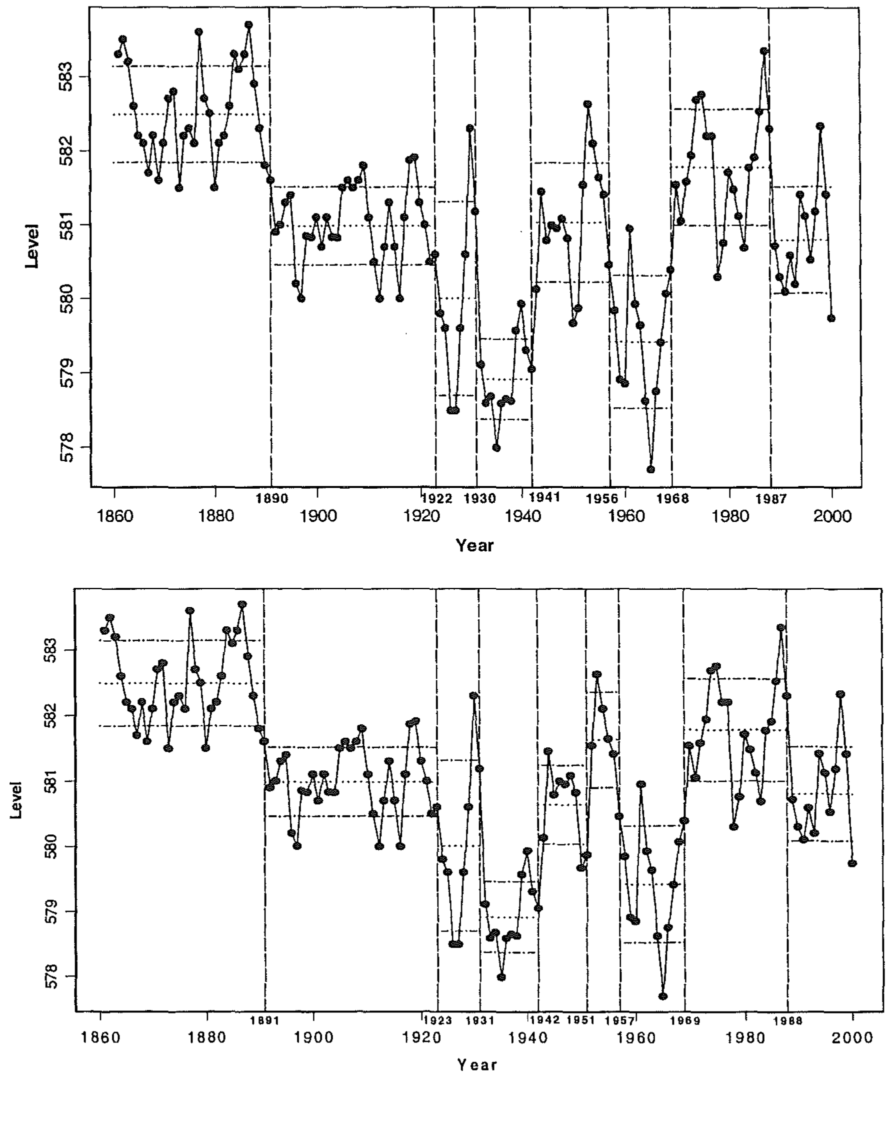

Figure 3: Final partitions for $G = 8, 9$. The horizontal lines indicate μ and $\mu \pm \sigma$.